PHIL 7001: Fundamentals of AI, Data, and Algorithms

Week 3
Probability

Boris Babic,
HKU 100 Associate Professor of Data Science, Law and Philosophy

Today

- Probability is a way of quantifying the uncertainty of events.
- To do this, we (ordinarily) need to think think about that event relative to some space of possibilities.
- Bayes' Rule, named after the Reverend Thomas Bayes, provides a way to update probabilities (or theories/hypotheses) on the basis of additional evidence.
- In data science, they provide the mathematical framework to handle uncertainty and make predictions.

Today

Learning goals

- Fundamentals of Probability
- Axioms and Rules of Probability
- Conditional Probability
- Independence and Dependence of Events
- Introduction to Bayes' Rule
- Practical Applications of Probability and Bayes' Rule in Data Science

Review

Last week

Review of last week

- Introduction to R
- Basic commands and functions in R
- Plotting with R (ggplot2)
- **Building a regression model in R**
- Thinking critically about models

Practice

- You have a dataset named `data_students` containing information about students' study hours and their corresponding exam scores:
- ```
  data_students <- data.frame(
    study_hours = c(2, 4, 3, 5, 7, 6, 8, 9, 10),
    exam_scores = c(60, 75, 68, 80, 90, 85, 92, 94, 98)
    )
  ```
- You want to visualize the relationship between study hours and exam scores using a scatter plot with a regression line. The x-axis should be labeled **Study Hours**, and the y-axis should be labeled **Exam Scores**.
- Please take 5-7 minutes to complete this on your own.
  ```

A Sample Exam Question

- You have a dataset named `data_students` containing information about students' study hours and their corresponding exam scores:
- ```
  data_students <- data.frame(
    study_hours = c(2, 4, 3, 5, 7, 6, 8, 9, 10),
    exam_scores = c(60, 75, 68, 80, 90, 85, 92, 94, 98)
    )
  ```
- You want to visualize the relationship between study hours and exam scores using a scatter plot with a regression line. The x-axis should be labeled **Study Hours**, and the y-axis should be labeled **Exam Scores**.
- Please fill in the blanks below.
- ```
  ggplot(_______, aes(x=_______, y=_______)) +
    geom_point() +
    geom_smooth(method = "lm", se = FALSE, color = "blue") +
    labs(x='_______', y='_______',
    title='Regression Model of Study Hours vs Exam Scores')
  ```

Probability

Probability

- Tyche: Goddess of chance (daughter of Zeus).
- The ancient Greeks believed that when no other cause can be attributed to random events such as floods, droughts, frosts, then Tyche is responsible.
- Probability is the study of such random events or, more generally, of randomness.

Probability

- But what is probability?
- For our purposes, the probability of an event can be interpreted as the long-run frequency of the occurrence of that event.
- Always???
- What is the probability of a nuclear war in the next year? Of US dollar collapse?
- When Galileo first observed Saturn through a telescope, he saw something like this.

- Are those rings around the planet? Handles? Or is it three planets next to each other? Can you assign a probability?

Probability vs. statistics

- **What is the difference between probability and statistics?**
- A probability question: A fair die will be tossed twice. What is the probability that it lands on six both times?
- A (descriptive) statistics question: A die was tossed twice and it landed on a five and a six. What is the mean die value?
- An (inferential) statistics question: A die was tossed twice and it landed on a five and a six. How confident are you that the die is fair?
- To answer this question, can you use descriptive statistics? Can you use probability? How? Is there a right answer? What is it?

Basic Concepts in Probability

- An **experiment**, in the literal sense, is an action or process that leads to one of several possible outcomes.
- In this course, we will use it in a general sense where we are doing a task and the outcomes of that task are random (what will be the outcome of the test is not known beforehand)
  - For example, tossing a coin, rolling a dice, picking a card from a deck, recording the stock prices, checking the blood pressure of a patient etc.
- **Outcome**: A possible result of a probability experiment.
- **Sample space** (denoted by $\boldsymbol{\Omega}$) represents the collection of all possible outcomes of an experiment
  - for tossing a coin, $\boldsymbol{\Omega} = \{\text{H, T}\}$
  - for rolling a dice, $\boldsymbol{\Omega} = \{1,2,3,\ldots,6\}$
- **Event**: Any subset of the sample space.
  - Let A represent an event where the outcome of a dice is even $- >$ A $= \{2,4,6\}$

Intuition behind Probability

- What does it mean to say the probability of event A is x?
  - Tossing a coin: the probability of {H} is 0.5
  - Weather forecast: there is a 30% chance of rain tomorrow
- The intuition is that if we observe today's conditions over and over again, 30% of the 'tomorrows' will result in rain.
- We often think of relative frequency (how many times something happens out of the total) as the probability of an event.
- However, this is not based on rigorous mathematical theory.

Random variables

- A random variable maps events to real numbers.
  For example, $\{H\} \to 0$ and $\{T\} \to 1$.
  This way, we can talk about the values 0 and 1 instead of using 'Heads' and 'Tails'.

- In probability we know which values our variables X can take, and we know how probable those values are.

- Ex: if the variable represents the outcome of the toss of a fair die (i.e., which face landed up), what are the values? How likely are they?

- Variables are
  - Discrete, corresponding to natural or counting numbers.
  - Continuous, corresponding to real numbers.

- Classify the following:
  - Height or weight
  - Number of monthly lottery winners in California
  - Temperature tomorrow
  - A parent's number of children
  - The amount of money in your bank account

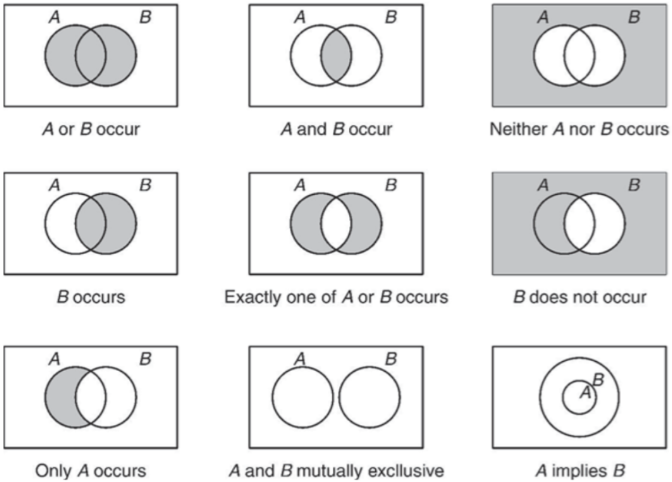
Union, Intersection, and Complement

- **Union ($\cup$)**: The union of two events A and B is the event that either A or B or both occur. It's denoted as $A \cup B$.
- **Intersection ($\cap$)**: The intersection of two events A and B is the event that both A and B occur. It's denoted as $A \cap B$.
- **Complement (^{c})**: The complement of an event A is the event that A does not occur. It's denoted as A' or A^c.
- **Disjoint or Mutually Exclusive Events**: If two events A and B cannot occur at the same time, we say they are disjoint or mutually exclusive. This means the intersection of A and B, denoted as $A \cap B$, is an empty set, denoted by $\varnothing$.
- **In Simple Terms**: If event A happens, event B cannot happen, and vice versa. They don't overlap at all. An example could be flipping a coin. The events "heads" and "tails" are mutually exclusive since you can't get both at the same time.

Rules of Probability and Useful Formulas

- A **Probability Function**, often denoted as P or Pr, assigns to each event A a number P(A). This number fulfills the following conditions:
  - $0 \leq P(A) \leq 1$.
  - The probability of the entire sample space (all possible outcomes) is 1: $P(\Omega) = 1$.
  - $P(A \cup B) = P(A) + P(B)$ when A and B are disjoint.
- **Useful Formulas**:
  - **Probability of a Union**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Examples:
  - Probability that a die lands on an even or odd number: $P(2, 4, 6) + P(1, 3, 5) = 0.5 + 0.5 = 1$.
  - Probability that a die lands on an even or prime number: $P(2, 4, 6) + P(2, 3, 5) - P(2) = 0.5 + 0.5 - 0.1 = 0.9$.
  - **Complementary Rule**: The probability of A not occurring (A^c) is: $P(A^c) = 1 - P(A)$.
  - **Product Rule**: For any two events A and B, $P(A \cap B) = P(A|B) * P(B)$
  - Example: Probability prime and even: P(2 given 2,4 or 6) $\times$ P(2,4 or 6) $= 1/3 \times 1/2 = 1/6$.
  - **Independent event**: Two events A and B are independent if the knowledge that one occurred does not affect the chance the other occurs. If A and B are independent, $P(A \cap B) = P(A) \times P(B)$.
  - : Example: Probability of A = 'Heads on Toss 1' and B = 'Heads on Toss 2' $= P(A) \times P(B) = 1/2 \times 1/2 = 1/4$.

Venn Diagram



Knowing your sample space

- A certain couple has two children. At least one of them is a boy. What is the probability that both children are boys?
- Possibilities: BB, BG, GB, GG
- What can we rule out? GG
- What remains: BB, BG, GB
- Probability that both children are boys is $1/3$.

Probability distributions

- A probability distribution is a function f of the random variable X: $f(X)$
- This function can only take values between 0 and 1.
- Also, this function is additive: for two independent events, the probability of their sum is the sum of their probabilities.
- Ex: Pr (die lands on 4) + Pr (die lands on 6) = Pr (die lands on 4 or die lands on 6).
- In the discrete case, it tells us how probable it is that the random variable X will take a specific value x. We often denote the function f with $\Pr$.
- Ex: For a fair die, $f(3) = \Pr(X = 3) = 1/6$.
- In the continuous case, it tells us how likely it is that our variable will be contained within an interval $[a, b]$.
- Ex: For a person's weight, $\Pr(a < x < b) = k$.
- But what about $\Pr(x)$ in the continuous case (???).

Exercise problem

- A fair die is rolled once. What is the probability that it lands on a number greater than 4? (easier)
- A fair die is rolled 5 times. What is the probability of seeing exactly the pattern 6, 5, 4, 3, followed by a 2 or a 1? (harder)

Exercise problem: Answers

- A fair die is rolled once. What is the probability that it lands on a number greater than 4?

$$\Pr(X > 4) = \Pr(X = 5) + \Pr(X = 6) = 2/6 = 1/3$$

- A fair die is rolled 5 times. What is the probability of seeing exactly the pattern 6, 5, 4, 3, followed by a 2 or a 1?

$$\left(\frac{1}{6}\right)^5 + \left(\frac{1}{6}\right)^5 = 0.0002$$

Bayes' Rule

Frequentist vs. Bayesian

- **Frequentist**: Probability is the long-run frequency of events. Data are considered random and parameters are fixed.
- **Bayesian**: Probability expresses a degree of belief in an event. Beliefs can be updated with data. Both data and parameters are considered random.
- In everyday language: Frequentists use probability only to model certain processes broadly described as "sampling," while Bayesians use probability more widely to model both sampling and other kinds of uncertainty.
- For Bayesians, uncertainty is epistemic. As a result, probability describes a certain kind of mental state within an agent.
- Philosophically, this is very different from frequentist probability.

Introduction to Bayesian Approach

- Imagine you are a doctor. A patient comes in and tells you they have a runny nose. You wonder: do they have a common cold, or is it an allergy? You can use Bayes' Rule to update your beliefs based on this new evidence (the runny nose).
- This is a classic example where Bayesian approaches are useful: we have prior knowledge (prevalence of colds and allergies), and we update our beliefs based on new evidence (runny nose).

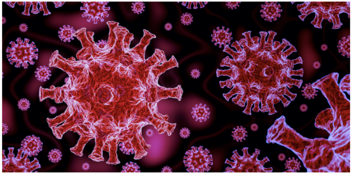
Bayes' Rule

- **Bayes' Rule**: A way to find a probability when we know certain other probabilities.
- **Formula**: $P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$
- **Prior (P(A))**: The initial degree of belief in A.
- **Likelihood (P(B|A))**: The probability of the evidence given that the hypothesis is true.
- **Marginal Likelihood (P(B))**: The total probability of the evidence.
- **Posterior (P(A|B))**: The updated belief after considering the evidence.

Bayes' Rule Example 1

- Suppose you learn that a dice landed on an even number. What is the probability that it landed on the number two, conditional on it landing on an even number?
- How to answer this question using Bayes' Rule?
- $P(B|A) = \Pr$ (even number | two)= 1 (this must be true)
- $P(A) = \Pr$ (two) = 1/6
- $P(B) = \Pr$ (even number) = 1/2
- Therefore: $[1*(1/6)] \ / \ (1/2) = (1/6)/(1/2) = (1/6)*2 = 1/3$.
- Intuition: There are three even numbers (2, 4, 6), and one of them is prime (2). Hence 1/3 is correct.

Bayes' Rule Example 2



- Consider a scenario where we are testing for lung cancer. The base rate (prevalence) of lung cancer in a certain population is 1%. Suppose that the sensitivity of the test is 99%, which means that 99% of the people who actually have lung cancer (True Positives) will test positive with this test. Additionally, the test has a specificity of 95%. This means that 95% of the people who do not have lung cancer (True Negatives) will test negative with this test.
- Now, imagine you get tested and the test comes back positive. What is the probability that you actually have the disease, given this positive result?
- This is a classic problem that can be solved using Bayes' Rule. Bayes' Rule allows us to update our belief about the probability of having the disease given a positive test result.

Bayes' Rule - Example Solution

- **Prior (P(D))**: Probability of having the disease = 0.01.
- **Likelihood (P(T|D))**: Probability of a positive test given disease = 0.99 (Test is 99% accurate).
- **False Positive (P(T | Not D))**: Probability of a positive test given no disease = 1 - P(N | Not D) = 1 - 0.95 = 0.05.
- **Marginal Likelihood (P(T))**: Total probability of the evidence, i.e., of a positive test = P(T|D)P(D) + P(T| Not D)P(Not D) = 0.99 * 0.01 + 0.05 * 0.99 = 0.0594.
  This is due to the law of total probability. For example,
  P(A) = P(A|B) + P(A| not B).
  please make sure you understand this!
- **Posterior (P(D|T))**: Probability of disease given a positive test using Bayes' Rule =

$$\frac{P(T|D) \cdot P(D)}{P(T)} = \frac{0.99 \cdot 0.01}{0.0594} \approx 0.167 \text{ or } 16.7\%.$$

- So, even with a positive test with 99% sensitivity and 95% specificity, there's only about a 16.7% chance you have lung cancer.
- This is because lung cancer is very rare. As a result, we have to combine our information about its rarity with our information derived from the test itself.

Bayes' Rule: What can we learn from the example?

- Notice the difference between $\Pr(D|T)$ (0.167) and $\Pr(T|D)$ (0.99)!
- High specificity and sensitivity can still lead to extremely low posterior probability
- A lesson to remember for AI and machine learning!

Bayes' Rule: Exercise

- Now, given the same context of testing for lung cancer, can you try calculating the probability of having the disease given a negative test result $P(D|N)$ using Bayes' Rule?
- Recall that the base rate (prevalence) of lung cancer is 1%, and that the test has 99% sensitivity and 95% specificity.

Bayes' Rule: Exercise

- Now, given the same context of testing for lung cancer, can you try calculating the probability of having the disease given a negative test result P(D|N) using Bayes' Rule?
- Recall that the base rate (prevalence) of lung cancer is 1%, and that the test has 99% sensitivity and 95% specificity.
- **Answer:**

$$P(D|N) = \frac{P(N|D) \cdot P(D)}{P(N)} = \frac{0.01 \cdot 0.01}{0.01 \cdot 0.01 + 0.95 \cdot 0.99} \approx 0.0001 \text{ or } 0.01\%$$

Application of Bayes' Rule in Data Science

- Predictive Modeling: For instance, in spam filtering, an email can be considered as spam or not spam.
- Medical diagnoses: Doctors can use it to determine the probability of a patient having a disease.
- Recommendation Systems: Based on the user's activity, the system recommends products.

Summary

Today

- Probability is a way of quantifying the uncertainty associated with events chosen from some universe of events.
- Bayes' Rule, named after the Reverend Thomas Bayes, provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence.
- In data science, it provides the mathematical framework to handle uncertainty and make predictions.

Coming Up

- Case Study 1: Monty Hall
- Inference: Decision, Estimation and Hypothesis Tests
- Bayesian Statistics